METHOD OF DESIGNING AGONISTS AND ANTAGONISTS TO IGF RECEPTOR

Field of the Invention

This invention relates to the field of receptor structure and receptor/ligand interactions. In particular it relates to the field of using receptor structure to predict the structure of related receptors and to the use of the determined structures and predicted structures to select and screen for agonists and antagonists of the polypeptide ligands.

Background of the Invention

Insulin is the peptide hormone that regulates glucose uptake and metabolism. The two types of diabetes mellitus are associated either with an inability to produce insulin because of destruction of the pancreatic islet cells (Homo-Delarche, F. & Boitard, C.,1996, Immunol. Today 10: 456-460) or with poor glucose metabolism resulting from either insulin resistance at the target tissues, or from inadequate insulin secretion by the islets or faulty liver function (Taylor, S. I., et al., 1994, Diabetes, 43: 735-740).

Insulin-like growth factors-1 and 2 (IGF-1 and 2) are structurally related to insulin, but are more important in tissue growth and development than in metabolism. They are primarily produced in the liver in response to growth hormone, but are also produced in most other tissues, where they function as paracrine/autocrine regulators. The IGFs are strong mitogens, and are involved in numerous physiological states and certain cancers (Baserga, R., 1996, TibTech 14: 150-152).

Epidermal growth factor (EGF) is a small polypeptide cytokine that is unrelated to the insulin/IGF family. It stimulates marked proliferation of epithelial tissues, and is a member of a larger family of structurally-related cytokines, such as transforming growth factor α, amphiregulin, betacellulin, heparin-binding EGF and some viral gene products. Abnormal EGF family signalling is a characteristic of certain cancers (Soler, C. & Carpenter, G., 1994 In Nicola, N. (ed) Guidebook to Cytokines and Their receptors", Oxford Univ. Press, Oxford, pp194-197; Walker, F. & Burgess, A. W., 1994, In Nicola, N. (ed) Guidebook to Cytokines and Their receptors", Oxford Univ. Press, Oxford, pp198-201).

Each of these growth factors mediates its biological actions through binding to the corresponding receptor. The IR, IGF-1R and the insulin receptor-related receptor (IRR), for which the ligand is not known, are closely related to each other, and are referred to as the insulin receptor subfamily. A

large body of information is now available concerning the primary structure of these insulin receptor subfamily members (Ebina, Y., et al., 1985 Cell 40: 747-758; Ullrich, A., et al., 1985, Nature 313: 756-761; Ullrich, A. et al., 1986, EMBO J 5: 2503-2512; Shier, P. & Watt, V. M., 1989, J. Biol. Chem. 264: 14605-14608) and the identification of some of their functional domains (for reviews see De Meyts, P. 1994, Diabetologia 37: 135-148; Lee, J. & Pilch, P. F. 1994 Amer. J. Physiol. 266: C319-C334.; Schaffer, L. 1994, Eur. J. Biochem. 221: 1127-1132). IGF-1R, IR and IRR are members of the tyrosine kinase receptor superfamily and are closely related to the epidermal growth factor receptor (EGFR) subfamily, with which they share significant sequence identity in the extracellular region as well as in the cytoplasmic kinase domains (Ullrich, A. et al., 1984 Nature 309: 418-425; Ward, C. W. et al., 1995 Proteins: Structure Function & Genetics 22: 141-153). Both the insulin and EGF receptor subfamilies have a similar arrangement of two homologous domains (L1 and L2) separated by a cys-rich region of approximately 160 amino acids containing 22-24 cys residues (Bajaj, M., et al., 1987 Biochim. Biophys. Acta 916: 220-226; Ward, C. W. et al., 1995 Proteins: Structure Function & Genetics 22: 141-153). The C-terminal portion of the IGF-1R ectodomain (residues 463 to 906) is comprised of four domains: a connecting domain, two fibronectin type 3 (Fn3) repeats, and an insert domain (O'Bryan, J. P., et al., 1991 Mol Cell Biol 11: 5016-5031). The C-terminal portion of the EGFR ectodomain (residues 477-621) consists solely of a second cys-rich region containing 20 cys residues (Ullrich, A. et al., 1984, Nature 309: 418-425).

Little is known about the secondary, tertiary and quaternary structure of the ectodomains of these receptor subfamilies. Unlike the members of the EGFR subfamily which are transmembrane monomers which dimerise on binding ligand, the IR subfamily members are homodimers, held together by disulphide bonds. The extracellular region of the IR/IGF-1R/IRR monomers contains an α-chain (~ 703 to 735 amino acid residues) and 192-196 residues of the ß-chain. There is a ~23 residue transmembrane segment, followed by the cytoplasmic portion (354 to 408 amino acids), which contains the catalytic tyrosine kinase domain flanked by juxtamembrane and C-tail regulatory regions and is responsible for mediating all receptor-specific functions (White, M. F. & Kahn, C. R. 1994 J. Biol. Chem. 269: 1-4). Chemical analyses of the receptor suggest that the α-chains are linked to the ß-chains

via a single disulphide bond, with the IR dimer being formed by at least two α-α disulphide linkages (Finn, F. M., et al., 1990, Proc. Natl. Acad. Sci. 87: 419-423; Chiacchia, K. B., 1991, Biochem. Biophys. Res. Commun. 176, 1178-1182; Schaffer, L. & Ljungqvist, L., 1992, Biochem. Biophys. Res. Comm. 189: 650-653; Sparrow, L. G., et al., 1997, J. Biol. Chem. 47: 29460-29467).

Although the three-dimensional (3D) structures of the ligands EGF, TGF-alpha (Hommel, U., et al., 1992, J. Mol. Biol. 227:271-282), insulin (Dodson, E. J., et al., 1983, Biopolymers 22:281-291), IGF-1 (Sato, A., et al., 1993, Int J Peptide Protein Res 41:433-440) and IGF-2 (Torres, A. M., et al.,1995, J. Mol. Biol. 248:385-401) are known, and numerous analytical and functional studies of ligand binding to EGFR (Soler, C. & Carpenter, G., 1994 In Nicola (ed) Guidebook to Cytokines and Their receptors", Oxford Univ. Press, Oxford, pp194-197), IGF-1R and IR (see De Meyts, P., 1994 Diabetologia, 37:135-148) have been carried out, the mechanisms of ligand binding and subsequent transmembrane signalling have not been resolved.

Ligand-induced, receptor-mediated phosphorylation is the signalling mechanism by which most cytokines, polypeptide hormones and membrane-anchored ligands exert their biological effects. The primary kinase may be part of the intracellular portion of the transmembrane receptor protein, as in the tyrosine kinase receptors (for review see Yarden, Y., et al., 1988, Ann. Rev. Biochem. 57:443-478) or the Ser/Thr kinase receptors (Alevizopoulos, A. & Mermod, N., 1997, BioEssays, 19:581-591) or may be non-covalently associated with the cytoplasmic tail of the transmembrane protein(s) making up the receptor complex, as in the case of the haemopoietic growth factor receptors (Stahl, N., et al., 1995, Science 267:1349-1353). The end result is the same, ligand binding leads to receptor dimerization or oligomerization or a conformational change in pre-existing receptor dimers or oligomers, resulting in activation by transphosphorylation, of the covalently attached or non-covalently associated protein kinase domains (Hunter, T., 1995, Cell, 80:225-236).

Many oncogenes have been shown to be homologous to growth factors, growth factor receptors or molecules in the signal transduction pathways (Baserga, R.,1994 Cell, 79:927-930; Hunter, T., 1997 Cell, 88:333-346). One of the best examples is v-Erb (related to the EGFR). Since overexpression of a number of growth factor receptors results in ligand-dependent transformation, an alternate strategy for oncogenes is to regulate

the expression of growth factor receptors or their ligands or to directly bind to the receptors to stimulate the same effect (Baserga, R., 1994 Cell, 79:927-930). Examples are v-Src, which activates IGF-1 R intracellularly; c-Myb, which transforms cells by enhancing the expression of IGF1R; and SV40 T antigen which interacts with the IGF-1R and enhances the secretion of IGF-1 (see Baserga, R.,1994 Cell, 79:927-930 for review). Cells in which the IGF-1R has been disrupted or deleted cannot be transformed by SV40 T antigen. If oncogenes activate growth factors and their receptors, then tumour suppressor genes should have the opposite effect. One good example of this is the Wilm's tumour suppressor gene, WT1, which suppresses the expression of IGF-1R (Drummond, J. A., et al., 1992, Science, 257:275-277). Cells that are driven to proliferate by oncogenes undergo massive apoptosis when growth factor receptors are ablated, since, unlike normal cells, they appear unable to withdraw from the cell-cycle and enter into the G_0 phase (Baserga, R.,1994 Cell, 79:927-930).

The insulin-like growth factor-1 receptor (IGF-1R) is one of several growth-factor receptors that regulate the proliferation of mammalian cells. However, its ubiquitousness and certain unique aspects of its function make IGF-1R an ideal target for specific therapeutic interventions against abnormal growth, with very little effect on normal cells (see Baserga, R., 1996 TIBTECH, 14:150-152). The receptor is activated by IGF1, IGF2 and insulin, and plays a major role in cellular proliferation in at least three ways: it is essential for optimal growth of cells *in vitro* and *in vivo*; several cell types require IGF-1R to maintain the transformed state; and activated IGF-1R has a protective effect against apoptotic cell death (Baserga, R., 1996 TIBTECH, 14:150-152). These properties alone make it an ideal target for therapeutic interventions. Transgenic experiments have shown that IGF-1R is not an absolute requirement for cell growth, but is essential for the establishment of the transformed state (Baserga, R.,1994 Cell, 79: 927-930). In several cases (human glioblastoma, human melanoma; human breast carcinoma; human lung carcinoma; human ovarian carcinoma; human rhabdomyosarcoma; mouse melanoma, mouse leukaemia; rat glioblastoma; rat rhabdomyosarcoma; hamster mesothelioma) the transformed phenotype can be reversed by decreasing the expression of IGF-1R using antisense to IGF-1R (Baserga, R., 1996 TIBTECH 14:150-152); or by interfering with its function by antibodies to IGF-1R (human breast carcinoma; human

rhabdomyosarcoma) or by dominant negatives of IGF-1R (rat glioblastoma; Baserga, R.,1996 TIBTECH 14:150-152).

Three effects are observed when the function of IGF-1R is impaired: tumour cells undergo massive apoptosis which results in inhibition of tumourogenesis; surviving tumour cells are eliminated by a specific immune response; and such a host response can cause a regression of an established wild-type tumour (Resnicoff, M., et al., 1995, Cancer Res. 54:2218-2222). These effects, plus the fact that interference with IGF-1R function has a limited effect on normal cells (partial inhibition of growth without apoptosis) makes IGF-1R a unique target for therapeutic interventions (Baserga, R., 1996 TIBTECH 14:150-152). In addition IGF-1R is downstream of many other growth factor receptors, which makes it an even more generalised target. The implication of these findings is that if the number of IGF-1Rs on cells can be decreased or their function antagonised, then tumours cease to grow and can be removed immunologically. These studies establish that IGF-1R antagonists will be extremely important therapeutically.

Many cancer cells have constitutively active EGFR (Sandgreen, E. P., et al., 1990, Cell, 61:1121-135; Karnes, W. E. J., et al., 1992, Gastroenterology, 102:474-485) or other EGFR family members (Hines, N. E.,1993, Semin. Cancer Biol. 4:19-26). Elevated levels of activated EGFR occur in bladder, breast, lung and brain tumours (Harris, A. L., et al., 1989, In Furth & Greaves (eds) The Molecular Diagnostics of human cancer. Cold Spring Harbor Lab. Press, CSH, NY, pp353-357). Antibodies to EGFR can inhibit ligand activation of EGFR (Sato, J. D., et al., 1983 Mol. Biol. Med. 1:511-529) and the growth of many epithelial cell lines (Aboud-Pirak E., et al., 1988, J. Natl Cancer Inst. 85:1327-1331). Patients receiving repeated doses of a humanised chimeric anti-EGFR monoclonal antibody showed signs of disease stabilization. The large doses required and the cost of production of humanised monoclonal antibody is likely to limit the application of this type of therapy. These findings indicate that the development of EGF antagonists will be attractive anticancer agents.

Summary of the Invention

The present inventors have now obtained 3D structural information concerning the insulin-like growth factor receptor (IGF-1R). This information can be used to predict the structure of related members of the insulin

receptor family and provides a rational basis for the development of ligands for specific therapeutic applications.

Accordingly, in a first aspect the present invention provides a method of designing a compound able to bind to a molecule of the insulin receptor family and to modulate an activity mediated by the molecule, including the step of assessing the stereochemical complementarity between the compound and the receptor site of the molecule, wherein the receptor site includes:

(a) amino acids 1 to 462 of the receptor for IGF-1, having the atomic coordinates substantially as shown in Figure 1;

(b) a subset of said amino acids, or;

(c) amino acids present in the amino acid sequence of a member of the insulin receptor family, which form an equivalent three-dimensional structure to that of the receptor molecule as depicted in Figure 1.

The phrase "insulin receptor family" encompasses, for example, IGF-1R, IR and IRR. In general, insulin receptor family members show similar domain arrangements and share significant sequence identity (preferably at least 40% identity).

By "stereochemical complementarity" we mean that the biologically active substance or a portion thereof correlates, in the manner of the classic "lock-and-key" visualisation of ligand-receptor interaction, with the groove in the receptor site.

In a preferred embodiment of this aspect of the invention, the compound is selected or modified from a known compound identified from a database.

In a further preferred embodiment, the compound is designed so as to complement the structure of the receptor molecule as depicted in Figure 1.

In a further preferred embodiment, the compound has structural regions able to make close contact with amino acid residues at the surface of the receptor site lining the groove, as depicted in Figure 2.

In a further preferred embodiment, the compound has a stereochemistry such that it can interact with both the L1 and L2 domains of the receptor site.

In a further preferred embodiment, the compound has a stereochemistry such that it can interact with the L1 domain of a first monomer of the receptor homodimer, and with the L2 domain of the other monomer of the receptor homodimer.

In a further preferred embodiment, the interaction of the compound with the receptor site alters the position of at least one of the L1, L2 or cysteine-

rich domains of the receptor molecule relative to the position of at least one of the other of said domains. Preferably, the compound interacts with the β sheet of the L1 domain of the receptor molecule, thereby causing an alteration in the position of the L1 domain relative to the position of the cysteine-rich domain or of the L2 domain. Alternatively, the compound interacts with the receptor site in the region of the interface between the L1 domain an the cysteine-rich domain of the receptor molecule, thereby causing the L1 domain and the cysteine-rich domain to move away from each other. In another preferred embodiment, the compound interacts with the hinge region between the L2 domain and the cysteine-rich domain of the receptor molecule, thereby causing an alteration in the positions of the L2 domain and the cysteine-rich domain relative to each other.

In a further preferred embodiment, the stereochemical complementarity between the compound and the receptor site is such that the compound has a K_b for the receptor side of less than 10^{-6}M, more preferably is less than 10^{-8}M.

In a further preferred embodiment or the first aspect of the present invention, the compound has the ability to increase an activity mediated by the receptor molecule.

In a further preferred embodiment, the compound has the ability to decrease an activity mediated by the receptor molecule. Preferably, the stereochemical interaction between the compound and the receptor site is adapted to prevent the binding of a natural ligand of the receptor molecule to the receptor site. It is preferred that the compound has a K_i of less than 10^{-6}M, more preferably less than 10^{-8}M and more preferably less than 10^{-9}M.

In a further preferred embodiment of the first aspect of the present invention, the receptor is the IGF-1R, or the insulin receptor.

In a second aspect, the present invention provides a computer-assisted method for identifying potential compounds able to bind to a molecule of the insulin receptor family and to modulate an activity mediated by the molecule, using a programmed computer including a processor, an input device, and an output device, including the steps of:

(a) inputting into the programmed computer, through the input device, data comprising the atomic coordinates of the IGF-1R molecule as shown in Figure 1, or a subset thereof;

(b) generating, using computer methods, a set of atomic coordinates of a structure that possesses stereochemical complementarity to the atomic

coordinates of the IGF-1R site as shown in Figure 1, or a subset thereof, thereby generating a criteria data set;

(c) comparing, using the processor, the criteria data set to a computer database of chemical structures;

(d) selecting from the database, using computer methods, chemical structures which are structurally similar to a portion of said criteria data set; and

(e) outputting, to the output device, the selected chemical structures which are similar to a portion of the criteria data set.

In a preferred embodiment of the second aspect, the programmed computer includes a data storage system which includes the dtatbase of chemical structures.

In a preferred embodiment of the second aspect, the method is used to identify potential compounds which have the ability to decrease an activity mediated by the receptor.

In another preferred embodiment, the computer-assisted method further includes the step of selecting one or more chemical structures from step (e) which interact with the receptor site of the molecule in a manner which prevents the binding of natural ligands to the receptor site.

In another preferred embodiment, the computer-assisted method further includes the step of obtaining a compound with a chemical structure selected in steps (d) and (e), and testing the compound for the ability to decrease an activity mediated by the receptor.

In a further preferred embodiment, the computer-assisted method is used to identify potential compounds which have the ability to increase an activity mediated by the receptor molecule.

In another preferred embodiment, the computer-assisted method further includes the step of obtaining a molecule with a chemical structure selected in steps (d) and (e), and testing the compound for the ability to increase an activity mediated by the receptor.

In a further preferred embodiment of the second aspect of the present invention, the receptor is the IGF-1R, or the insulin receptor.

In a third aspect, the present invention provides a method of screening of a putative compound having the ability to modulate the activity of a receptor of the insulin receptor family, including the steps of identifying a putative compound by a method according to the first or second aspects, and testing the

compound for the ability to increase or decrease an activity mediated by the receptor.

In a preferred embodiment of the third aspect, the test is carried out *in vitro*.

In a further preferred embodiment of the third aspect, the test is a high throughput assay.

In a preferred embodiment of the third aspect, the test is carried out *in vivo*.

Brief Description of the Drawings

Figure 1. IGF-1R residues 1-462, in terms of atomic coordinates refined to a resolution of 2.6 Å (average accuracy ≈ 0.3Å). The coordinates are in relation to a Cartesian system of orthogonal axes.

Figure 2. Depiction of the residues lining the groove of the IGF-1R receptor fragment 1-462.

Figure 3. Gel filtration chromatography of affinity-purified IGF-1R/462 protein. The protein was purified on a Superdex S200 column (Pharmacia) fitted to a BioLogic L.C. system (Biorad), equilibrated and eluted at 0.8 ml/min with 40 mM Tris/150 mM NaCl/0.02% NaN3 adjusted to pH 8.0. (a) Protein eluting in peak 1 contained aggregated IGF-1R/462 protein, peak 2 contained monomeric protein and peak 3 contained the c-myc undecapeptide used for elution from the Mab 9E10 immunoaffinity column. (b) Non-reduced SDS-PAGE of fraction 2 from IGF-1R/462 obtained following Superdex S200 (Fig.1a). Standard proteins are indicated.

Figure 4. Ion exchange chromatography of affinity-purified, truncated IGF-1R ectodomain. A mixture of gradient and isocratic elution chromatography was performed on a Resource Q column (Pharmacia) fitted to a BioLogic System (Biorad), using 20 mM Tris/pH 8.0 as buffer A and the same buffer containing 1M NaCl as buffer B. Protein solution in TBSA was diluted at least 1:2 with water and loaded onto the column at 2 ml/min. Elution was monitored by absorbance (280 nm) and conductivity (mS/cm). Target protein (peak 2) eluted isocratically with 20 mM Tris/0.14 M NaCl pH 8.0. Inset:

Isoelectric focusing gel (pH 3 - 7; Novex Australia Pty Ltd)of fraction 2. The pI was estimated at 5.1 from standard proteins (not shown).

Figure 5. Polypeptide fold for residues 1-462 of IGF-1R. The L1 domain is at the top, viewed from the N-terminal end and L2 is at the bottom. The space at the centre is of sufficient size to accommodate IGF-1. Helices are indicated by curled ribbon and b-strands by arrows. Cysteine side chains are drawn as ball-and-stick with lines showing disulfide bonds. The arrow points in the direction of view for L1 in Figure 7.

Figure 6. Amino acid sequences of IGF-1R and related proteins. a, L1 and L2 domains of human IGF-1R and IR are shown based on a sequence alignment for the two proteins and a structural alignment for the L1 and L2domains. Positions showing conservation physico-chemical properties of amino acids are boxed, residues used in the structural alignment are shown in Times Italic and residues which form the Trp 176 pocket are in Times Bold. Secondary structure elements for L1 (above the sequences) and L2 (below) are indicated as cylinders for helices and arrows for β-strands. Strands are shaded (pale, medium and dark grey) according to the β-sheet to which they belong. Disulfide bonds are also indicated. b, Cys-rich domains of human IGF-1R, IR and EGFR (domains 2 and 4) are aligned based on sequence and structural considerations. Secondary structural elements and disulfide bonds are indicated above the sequences. The dashed bond is only present in IR. Different types of disulfide bonded modules are labelled below the sequences as open, filled or broken lines. Boxed residues show conservation of physico-chemical properties and structurally conserved residues for modules 4-7 are shown in Times Italic. Residues from EGFR which do not conform to the pattern are in lowercase with probable disulfide bonding indicated below and the conserved Trp 176 and the semi-conserved Gln 182 are in Times Bold.

Figure 7. Stereo view of a superposition of the L1 (white) and L2 (black) domains. Residues numbers above are for L1 and below for L2. The side chain of Trp 176 which protrudes into the core of L1 is drawn as ball-and-stick.

Figure 8. Schematic diagram showing the association of three β-finger motifs. β-strands are drawn as arrows and disulfide bonds as zigzags.

Figure 9: Sequence alignment of hIGF-1R, hIR and hIRR ectodomains, derived by use of the PileUp program in the software package of the Genetics Computer Group, 575 Science Drive, Madison, Wisconsin, USA.
For assignment of homologous 3D structures see Figure 6.

Figure 10 Gel filtration chromatography of insulin receptor ectodomain and MFab complexes. hIR -11 ectodomain dimer (5 - 20 mg) was complexed with MFab derivatives (15-25 mg each) of the anti-hIR antibodies 18-44, 83-7 and 83-14 (Soos et al., 1986). Flution profiles were generated from samples loaded on to a Superdex S200 column (Pharmacia), connected to a BioLogic chromatography system (Biorad) and monitored at 280 nm. The column was eluted at 0.8 ml/min with 40 mM Tris/150 mM sodium chloride/0.02% sodium azide buffer adjusted to pH 8.0: Profile 0, hIR -11 ectodomain, Profile 1, ectodomain mixed with MFab 18-44; Profile 2 , ectodomain mixed with MFab18-44 and MFab 83-14; Profile 3, ectodomain mixed with MFab 18-44, MFab 83-14 and MFab 83-7. The apparent mass of each complex was determined from a plot of the following standard proteins: thyroglobulin (660 kDa), ferritin (440 kDa), bovine gamma globulin (158 kDa), bovine serum albumin (67 kDa), chicken ovalbumin (44 kDa) and equine myoglobin (17 kDa).

Figure 11 Schematic representations of electron microscopy images of the hIR ectodomain dimer.

Detailed Description of the Invention

We describe herein the expression, purification, and crystallization of a recombinant truncated IGF-1R fragment (residues 1-462) containing the L1-cysteine-rich-L2 region of the ectodomain. The selected truncation position is just downstream of the exon 6/exon 7 junction (Abbott. A. M., et al., 1992. J Biol Chem., 267:10759-10763), and occurs at a position where the sequences of the IR and EGFR families diverge markedly (Ward, C. W., et al.,1995, Proteins: Struct., Funct., Genet. 22:141-153; Lax, I., et al., 1988, Molec. Cellul. Biol. 8:1970-1978) suggesting it represents a domain boundary. To

limit the effects of glycosylation, the IGF-1R fragment was expressed in Lec8 cells, a glycosylation mutant of Chinese hamster ovary (CHO) cells, whose defined glycosylation defect produces N-linked oligosaccharides truncated at N-acetyl glucosamine residues distal to mannose residues (Stanley, P. 1989, Molec. Cellul. Biol. 9:377-383). Such an approach has facilitated glycoprotein crystallization (Davis, S. J., et al., 1993, Protein Eng. 6:229-232; Liu, J., et al., 1996, J. Biol. Chem. 271:33639-33646).

The IGF-1R construct described herein includes a c-myc peptide tag (Hoogenboom, H. R., et al.,1991, Nucleic Acids Res. 19:4133-4137) that is recognised by the Mab 9E10 (Evan, G. I., et al., 1985, Mol. Cell. Biol. 5:3610-3616) enabling the expressed product to be purified by peptide elution from an antibody affinity column followed by gel filtration over Superdex S200. The purified proteins crystallized under a sparse matrix screen (Jancarik, J. & Kim, S.-H., 1991, J. Appl. Cryst. 24:409-411) but the crystals were of variable quality, with the best diffracting to 3.0-3.5Å. Isocratic gradient elution by anion-exchange chromatography yielded protein that was less heterogenous and gave crystals of sufficient quality to determine the structure of the first three domains of the human IGF-1R.

The IGF-1R fragment consisted of residues 1-462 of IGF-1R linked via an enterokinase-cleavable pentapeptide sequence to an eleven residue c-myc peptide tag at the C-terminal end. The fragment was expressed in Lec8 cells by continuous media perfusion in a bioreactor using porous carrier disks. It was secreted into the culture medium and purified by peptide elution from an anti-c-myc antibody column followed by Superdex S200 gel filtration. The receptor fragment bound two anti-IGF-1R monoclonal antibodies, 24-31 and 24-60, which recognize conformational epitopes, but could not be shown to bind IGF-1 or IGF-2. Crystals of variable quality were grown as rhombic prisms in 1.7 M ammonium sulfate at pH 7.5 with the best diffracting to 3.0-3.5 Å. Further purification by isocratic elution on an anion-exchange column gave protein which produced better quality crystals, diffracting to 2.6 Å, that were suitable for X-ray structure determination.

The structure of this fragment (IGF-1R residues 1-462; L1-cys rich-L2 domains) has been determined to 2.6 Å resolution by X-ray diffraction. The L domains each adopt a compact shape consisting of a single stranded right-handed β-helix. The cys-rich region is composed of eight disulphide-bonded modules, seven of which form a rod-shaped domain with modules associated

in a novel manner. At the centre of this reasonably extended structure is a space, bounded by all three domains, and of sufficient size to accommodate a ligand molecule. Functional studies on IGF-1R and other members of the insulin receptor family show that the regions primarily responsible for hormone-binding map to this central site. Thus this structure gives a first view of how members of the insulin receptor family might interact with their ligands.

Another group has reported the crystallization of a related receptor, the EGFR, in a complex with its ligand EGF (Weber, W., et al., 1994, J Chromat. 679:181-189). However, difficulties were encountered with these crystals which diffracted to only 6 Å, insufficient for the determination of an atomic resolution structure of this complex (Weber, W., et al., 1994, J Chromat 679:181-189) or the generation of accurate models of structurally related receptor domains such as IGF-1R and IR by homology modelling.

The present inventors have developed 3D structural information about cytokine receptors in order to enable a more accurate understanding of how the binding of ligand leads to signal transduction. Such information provides a rational basis for the development of ligands for specific therapeutic applications, something that heretofore could not have been predicted *de novo* from available sequence data.

The precise mechanisms underlying the binding of agonists and antagonists to the IGF-1R site are not fully clarified. However, the binding of ligands to the receptor site, preferably with an affinity in the order of 10^{-8}M or higher, is understood to arise from enhanced stereochemical complementarity relative to naturally occurring IGF-1 ligands.

Such stereochemical complementarity, pursuant to the present invention, is characteristic of a molecule that matches intra-site surface residues lining the groove of the receptor site as eneumerated by the coordinates set out in Figure 1. The residues lining the groove are depicted in Figure 2. By "match" we mean that the identified portions interact with the surface residues, for example, via hydrogen bonding or by enthalpy-reducing Van der Waals interactions which promote desolvation of the biologically active substance within the site, in such a way that retention of the biologically active substance within the groove is favoured energetically.

Substances which are complemetary to the shape of the receptor site characterised by amino acids positioned at atomic coordinates set out in

Figure 1 may be able to bind to the receptor site and, when the binding is sufficiently strong, substantially prohibit binding of the naturally occurring ligands to the site.

It will be appreciated that it is not necessary that the complementarity between ligands and the receptor site extend over all residues lining the groove in order to inhibit binding of the natural ligand. Accordingly, agonists or antagonists which bind to a portion of the residues lining the groove are encompassed by the present invention.

In general, the design of a molecule possessing stereochemical complementarity can be accomplished by means of techniques that optimize, either chemically or geometrically, the "fit" between a molecule and a target receptor. Known techniques of this sort are reviewed by Sheridan and Venkataraghavan, Acc. Chem Res. 1987 20 322; Goodford, J. Med. Chem. 1984 27 557; Beddell, Chem. Soc. Reviews 1985, 279; Hol, Angew. Chem. 1986 25 767 and Verlinde C.L.M.J & Hol, W.G.J. Structure 1994, 2, 577, the respective contents of which are hereby incorporated by reference. See also Blundell et al., Nature 1987 326 347 (drug development based on information regarding receptor structure).

Thus, there are two preferred approaches to designing a molecule, according to the present invention, that complements the shape of IGF-1R or a related receptor molecule. By the geometric approach, the number of internal degrees of freedom (and the corresponding local minima in the molecular conformation space) is reduced by considering only the geometric (hard-sphere) interactions of two rigid bodies, where one body (the active site) contains "pockets" or "grooves" that form binding sites for the second body (the complementing molecule, as ligand). The second preferred approach entails an assessment of the interaction of respective chemical groups ("probes") with the active site at sample positions within and around the site, resulting in an array of energy values from which three-dimensional contour surfaces at selected energy levels can be generated.

The geometric approach is illustrated by Kuntz et al., J. Mol. Biol. 1982 161 269, the contents of which are hereby incorporated by reference, whose algorithm for ligand design is implemented in a commercial software package distributed by the Regents of the University of California and further described in a document, provided by the distributor, which is entitled "Overview of the DOCK Package, Version 1.0,", the contents of which are

hereby incorporated by reference. Pursuant to the Kuntz algorithm, the shape of the cavity represented by the IGF-R1 site is defined as a series of overlapping spheres of different radii. One or more extant data bases of crystallographic data, such as the Cambridge Structural Database System maintained by Cambridge University (University Chemical Laboratory, Lensfield Road, Cambridge CB2 1EW, U.K.) and the Protein Data Bank maintained by Brookhaven National Laboratory (Chemistry Dept. Upton, NY 11973, U.S.A.), is then searched for molecules which approximate the shape thus defined.

Molecules identified in this way, on the basis of geometric parameters, can then be modified to satisfy criteria associated with chemical complementarity, such as hydrogen bonding, ionic interactions and Van der Waals interactions.

The chemical-probe approach to ligand design is described, for example, by Goodford, J. Med. Chem. 1985 28 849, the contents of which are hereby incorporated by reference, and is implemented in several commercial software packages, such as GRID (product of Molecular Discovery Ltd., West Way House, Elms Parade, Oxford OX2 9LL, U.K.). Pursuant to this approach, the chemical prerequisites for a site-complementing molecule are identified at the outset, by probing the active site (as represented via the atomic coordinates shown in Fig. 1) with different chemical probes, e.g., water, a methyl group, an amine nitrogen, a carboxyl oxygen, and a hydroxyl. Favored sites for interaction between the active site and each probe are thus determined, and from the resulting three-dimensional pattern of such sites a putative complementary molecule can be generated.

The chemical-probe approach is especially useful in defining variants of a molecule known to bind the target receptor. Accordingly, crystallographic analysis of IGF-1 bound to the receptor site is expected to provide useful information regarding the interaction between the archetype ligand and the active site of interest.

Programs suitable for searching three-dimensional databases to identify molecules bearing a desired pharmacophore include: MACCS-3D and ISIS/3D (Molecular Design Ltd., San Leandro, CA), ChemDBS-3D (Chemical Design Ltd., Oxford, U.K.), and Sybyl/3DB Unity (Tripos Associates, St. Louis, MO).

Programs suitable for pharmacophore selection and design include: DISCO (Abbott Laboratories, Abbott Park, IL), Catalyst (Bio-CAD Corp., Mountain View, CA), and ChemDBS-3D (Chemical Design Ltd., Oxford, U.K.).

Databases of chemical structures are available from a number of sources including Cambridge Crystallographic Data Centre (Cambridge, U.K.) and Chemical Abstracts Service (Columbus, OH).

De novo design programs include Ludi (Biosym Technologies Inc., San Diego, CA), Sybyl (Tripos Associates) and Aladdin (Daylight Chemical Information Systems, Irvine, CA).

Those skilled in the art will recognize that the design of a mimetic may require slight structural alteration or adjustment of a chemical structure designed or identified using the methods of the invention.

The invention may be implemented in hardware or software, or a combination of both. However, preferably, the invention is implemented in computer programs executing on programmable computers each comprising a processor, a data storage system (including volatile and non-volatile memory and/or storage elements), at least one input device, and at least one output device. Program code is applied to input data to perform the functions described above and generate output information. The output information is applied to one or more output devices, in known fashion. The computer may be, for example, a personal computer, microcomputer, or workstation of conventional design.

Each program is preferably implemented in a high level procedural or object-oriented programming language to communicate with a computer system. However, the programs can be implemented in assembly or machine language, if desired. In any case, the language may be compiled or interpreted language.

Each such computer program is preferably stored on a storage medium or device (e.g., ROM or magnetic diskette) readable by a general or special purpose programmable computer, for configuring and operating the computer when the storage media or device is read by the computer to perform the procedures described herein. The inventive system may also be considered to be implemented as a computer-readable storage medium, configured with a computer program, where the storage medium so

configured causes a computer to operate in a specific and predefined manner to perform the functions described herein.

Compounds designed according to the methods of the present invention may be assessed by a number of *in vitro* and *in vivo* assays of hormone function. For example, the identification of IGF-1R antagonists of may be undertaken using a solid-phase receptor binding assay. Potential antagonists may be screened for their ability to inhibit the binding of europium-labelled IGF ligands to soluble, recombinant IGF-1R in a microplate-based format. Europium is a lanthanide fluorophore, the presence of which can be measured using time-resolved fluorometry. The sensitivity of this assay matches that achieved by radioisotopes, measurement is rapid and is performed in a microplate format to allow high-sample throughput, and the approach is gaining wide acceptance as the method of choice in the development of screens for receptor agonists/antagonists (see Apell et.al. J. Biomolec. Screening 3:19-27, 1998 : Inglese et. al. Biochemistry 37:2372-2377, 1998).

Binding affinity and inhibitor potency may be measured for candidate inhibitors using biosensor technology.

The IGF-1R antagonists may be tested for their ability to modulate receptor activity using a cell-based assay incorporating a stably transfected, IGF-1-responsive reporter gene [Souriau, C., Fort, P., Roux, P., Hartley, O., LeFranc, M-P. and Weill, M., 1997, Nucleic Acids Res. 25, 1585-1590]. An IGF-1-responsive, luciferase reporter gene has been assembled and transfected in 293 cells. The assay addresses the ability of IGF-1 to activate the reporter gene in the presence of novel ligands. It offers a rapid (results within 6-8 hours of hormone exposure), high-throughput (assay can be conducted in a 96-well format for automated counting) analysis using an extremely sensitive detection system (chemiluminescence). Once candidate compounds have been identified, their ability to antagonise signal transduction via the IGF-1R can be assessed using a number of routine *in vitro* cellular assays such as inhibition of IGF-1-mediated cell proliferation, induction of apoptosis in the presence of IGF-1 and the ablation of IGF-1-driven anchorage-independent cell growth in soft agar [D'Ambrosio, C., Ferber, A., Resnicoff, M. and Baserga, R., 1996, Cancer Res. 56, 4013-4020]. Such assays may be conducted on the P6 cell line, a cell line highly responsive to IGF as a result of the constitutive overexpression of the IGF-1R

(45-50,000 receptors/cell, [Pietrzkowski, Z., Sell, C., Lammers, R., Ullrich, A. and Baserga, R.,1992, Cell Growth.Diff. 3, 199-205]). Ultimately, the efficacy of any antagonist as a tumour therapeutic may be tested *in vivo* in animals bearing tumour isografts and xenografts as described [Resnicoff, M., Burgaud, J-L., Rotman, H. L., Abraham, D. and Baserga, R., 1995, Cancer Res. 55, 3739-3741; Resnicoff, M., Sell, C., Rubini, M., Coppola, D., Ambrose, D., Baserga, R. and Rubin, R., 1994 Cancer Res. 54: 2218-2222].

Tumour growth inhibition assays may be designed around a nude mouse xenograft model using a range of cell lines. The effects of the receptor antagonists and inhibitors may be tested on the growth of subcutaneous tumours.

A further use of the structure of the IGF-1R fragment described here is in facilitating structure determination of a related protein, such as a larger fragment of this receptor, another member of the insulin receptor family or a member of the EGF receptor family. This new structure may be either of the protein alone, or in complex with its ligand. For crystallographic analysis this is achieved using the method of molecular replacement (Brunger, Meth. Enzym. 1997 276 558-580, Navaza and Saludjian, *ibid.* 581-594, Tong and Rossmann, *ibid.* 594-611, Bentley, *ibid.* 611-619) in a program such as XPLOR. In this procedure diffraction data is collected from a crystalline protein of unknown structure. A transform of these data (Patterson function) is compared with a Patterson function calculated from a known structure. Firstly, the one Patterson function is rotated on the other to determine the correct orientation of the unknown molecule in the crystal. The translation function is then calculated to determine the location of the molecule with respect to the crystal axes. Once the molecule has been correctly positioned in the unit cell initial phases for the experimental data may be calculated. These phases are necessary for calculation of an electron density map from which structural differences may be observed and for refinement of the structure. Due to limitations in the method the search molecule must be structurally related to that which is to be determined. However it is sufficient for only part of the unknown structure (e.g. < 50%) to be similar to the search molecule. Thus the three dimensional structure of IGF-1R residues 1-462 may be used to solve structures consisting of related receptors, enabling a program of drug design as outlined above.

In summary, the general principles of receptor-based drug design can be applied by persons skilled in the art, using the crystallographic results presented above, to produce ligands of IGF-1R or other related receptors, having sufficient stereochemical complementarity to exhibit high affinity binding to the receptor site.

The present invention is further described below with reference to the following, non-limiting examples.

EXAMPLE 1

Expression, Purification and Crystallization of the IGF-1R Fragment.

Several factors hamper macromolecular crystallization including sample selection, purity, stability, solubility (McPherson, A., et al., 1995, Structure 3:759-768); Gilliland, G. L., & Ladner, J. E., 1996, Curr. Opin. Struct. Biol. 6:595-603), and the nature and extent of glycosylation (Davis, S. J., et al., 1993, Protein Eng. 6:229-232). Initial attempts to obtain structural data from soluble IGF-1R ectodomain (residues 1-906) protein, expressed in Lec8 cells (Stanley, P. 1989, Molec. Cellul. Biol. 9:377-383) and purified by affinity chromatography, produced large, well-formed crystals (1.0 mm x 0.2 mm x 0.2 mm) which gave no discernible X-ray diffraction pattern (unpublished data). Similar difficulties have been encountered with crystals of the structurally-related epidermal growth factor receptor (EGFR) ectodomain, which diffracted to only 6 Å, insufficient for the determination of an atomic resolution structure (Weber, W. et al., 1994, J Chromat 679:181-189). This prompted us to search for a fragment of IGF-1R that was more amenable to X-ray crystallographic studies.

The fragment expressed (residues 1-462) comprises the L1-cysteine-rich-L2 region of the ectodomain. The selected truncation position at Val462 is four residues downstream of the exon 6/exon 7 junction (Abbott, A. M., et al., 1992, J Biol Chem. 267:10759-10763), and occurs at a position where the sequences of the IR and the structurally related EGFR families diverge markedly (Lax, I., et al., 1988, Molec Cell Biol. 8:1970-1978; Ward, C. W., et al., 1995, Proteins: Struct., Funct., Genet. 22:141-153), suggesting that it represents a domain boundary. The expression strategy included use of the pEE14 vector (Bebbington, C. R. & Hentschel, C. C. G., 1987, In: Glover, D. M., ed. DNA Cloning. Academic Press, San Diego. Vol 3, p163) in glycosidase-defective Lec8 cells (Stanley, P., 1989, Molec. Cellul. Biol. 9:377-

383), which produce N-linked oligosaccharides lacking the terminal galactose and N-acetylneuraminic acid residues (Davis, S. J., et al., 1993, Protein Eng. 6:229-232; Liu, T., et al., 1996, J Biol Chem 271:33639-33646.). The construct contained a C-terminal c-myc affinity tag (Hoogenboom, H. R., et al., 1991, Nucl Acids Res. 19:4133-4137), which facilitated immunoaffinity purification by specific peptide elution and avoided aggressive purification conditions. These procedures yielded protein which readily crystallized after a further gel filtration purification step. This provided a general protocol to enhance crystallisation prospects for labile, multidomain glycoproteins.

The structure of this fragment is of considerable interest, since it contains the major determinants governing insulin and IGF-1 binding specificity (Gustafson, T. A. & Rutter, W. J., 1990, J. Biol. Chem. 265:18663-18667; Andersen, A. S., et al., 1990, Biochemistry, 29:7363-7366; Schumacher, R., et al., 1991, J. Biol. Chem. 266:19288-19295; Schumacher, R., et al., 1993, J. Biol. Chem. 268:1087-1094; Schäffer, L., et al., 1993, J. Biol. Chem. 268:3044-3047; Williams, P. F., et al., 1995, , J. Biol. Chem. 270:3012-3016), and is very similar to an IGF-1R fragment (residues 1-486) reported to act as a strong dominant negative for several growth functions and which induces apoptosis of tumour cells *in vivo* (D'Ambrosio, C., et al., 1996, Cancer Res. 56:4013-4020).

The expression plasmid pEE14/IGF-1R/462 was constructed by inserting the oligonucleotide cassette:

AatII

5' GACGTC GACGATGACGATAAG GAACAAAAACTCATC

D V D D D D K E Q K L I

(EK cleavage) (c-myc tail)

S E E D L N (Stop)

TCAGAAGAGGATCTGAAT TAGAATTC GACGTC 3'

EcoRI *AatII*

encoding an enterokinase cleavage site, c-myc epitope tag (Hoogenboom, H. R., et al., 1991, Nucleic acids Res. 19:4133-4137) and stop codon into the AatII site (within codon 462) of Igf-1r cDNA in the mammalian expression vector pECE (Ebina, Y., et al., 1985, Cell, 40:747-758; kindly supplied by W. J. Rutter, UCSF, USA), and introducing the DNA comprising the 5' 1521 bp of

the cDNA (Ullrich, A., et al., 1986, EMBO J. 5:2503-2512) ligated to the oligonucleotide cassette into the EcoRI site of the mammalian plasmid expression vector pEE14 (Bebbington, C. R. & Hentschel, C. C. G., 1987, In: Glover, D. M., ed. DNA Cloning. Academic Press, San Diego. Vol 3, p163; Celltech Ltd., UK). Plasmid pEE14/IGF-1R/462 was transfected into Lec8 mutant CHO cells (Stanley, P. 1989, Molec. Cellul. Biol. 9:377-383) obtained from the American Tissue Culture Collection (CRL:1737), using Lipofectin (Gibco-BRL). Cell lines were maintained after transfection in glutamine-free medium (Glascow modification of Eagle's medium (GMEM; ICN Biomedicals, Australia) and 10% dialysed FCS (Sigma, Australia) containing 25 μM methionine sulphoximine (MSX; Sigma, Australia) as described (Bebbington, C. R. & Hentschel, C. C. G., 1987, In: Glover, D. M., ed. DNA Cloning. Academic Press, San Diego. Vol 3, p163). Transfectants were screened for protein expression by Western blotting and sandwich enzyme-linked immunosorbent assay (ELISA) (Cosgrove, L., et al., 1995,) using monoclonal antibody (Mab) 9E10 (Evan et al., 1985) as the capture antibody, and either biotinylated anti-IGF-1R Mab 24-60 or 24-31 for detection(Soos et al., 1992; gifts from Ken Siddle, University of Cambridge, UK). Large-scale cultivation of selected clones expressing IGF-1R/462 was carried out in a Celligen Plus bioreactor (New Brunswick Scientific, USA) containing 70 g Fibra-Cel Disks (Sterilin, UK) as carriers in a 1.25 L working volume. Continuous perfusion culture using GMEM medium supplemented with non-essential amino acids, nucleosides, 25 μM MSX and 10% FCS was maintained for 1 to 2 weeks followed by the more enriched DMEM/F12 without glutamine, with the same supplemention for the next 4-5 weeks. The fermentation production run was carried out three times under similar conditions, and resulted in an estimated overall yield of 50 mg of receptor protein from 430 L of harvested medium. Cell growth was poor during the initial stages of the fermentation when GMEM medium was employed, but improved dramatically following the switch to the more enriched medium. Target protein productivity was essentially constant during the period from ~100 to 700 h of the 760 h fermentation, as measured by ELISA using Mab 9E10 as the capture antibody and biotinylated Mab 24-31 as the developing antibody.

Soluble IGF-1R/462 protein was recovered from harvested fermentation medium by affinity chromatography on columns prepared by coupling Mab 9E10 to divinyl sulphone-activated agarose beads (Mini Leak;

Kem En Tec. Denmark) as recommended by the manufacturer. Mini-Leak Low and Medium affinity columns with antibody loadings of 1.5-4.5 mg/ml of hydrated matrix were obtained, with the loading range of 2.5-3 mg/ml giving optimal performance (data not shown). Mab 9E10 was produced by growing hybridoma cells (American Tissue Culture Collection) in serum-free medium in the Celligen Plus bioreactor and recovering the secreted antibody (4 g) using protein A glass beads (Prosep-A, Bioprocessing Limited, USA). Harvested culture medium containing IGF-1R/462 protein was adjusted to pH 8.0 with Tris-HCl (Sigma), made 0.02% (w/v) in sodium azide and passed at 3-5 ml/min over 50 ml Mab 9E10 antibody columns at 4° C. Bound protein was recovered by recycling a solution of 2-10 mg of the undecamer c-myc peptide EQKLISEEDLN (Hoogenboom et al., 1991) in 20 ml of Tris-buffered saline containing 0.02% sodium azide (TBSA). Between 65% and 75% of the product was recovered from the medium as estimated by ELISA, with a further 15-25% being recovered by a second pass over the columns. Peptide recirculation (~10 times) through the column eluted bound protein more efficiently than a single, slower elution. Residual bound protein was eluted with sodium citrate buffer at pH 3.0 into 1 M Tris HCl pH 8.0 to neutralize the eluant, and columns were re-equilibrated with TBSA.

Gel filtration over Superdex S200 (Pharmacia, Sweden), of affinity-purified material showed a dominant protein peak at ~63 kDa, together with a smaller quantity of aggregated protein (Figure 3a). The peak protein migrated primarily as two closely spaced bands on reduced , sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE; Figure 3b), reacted positively in the ELISA with both Mab 24-60 and Mab 24-31, and gave a single sequence corresponding to the N-terminal 14 residues of IGF-1R. No binding of IGF-1 or IGF-2 could be detected in the solid plate binding assay (Cosgrove et al., 1995, Protein Express Purif. 6:789-798). The IGF-1R/462 fragment was further purified by ion-exchange chromatography on Resource Q (Pharmacia, Sweden). Using shallow salt gradients, protein enriched in the slowest migrating SDS-PAGE band was obtained (data not shown), which formed relatively large, well-formed crystals (see below). Isoelectric focussing showed the presence of one major and two minor isoforms. Protein purified on Resource Q with an isocratic elution step of 0.14 M NaCl in 20 mM TrisCl at pH 8.0 (fraction 2, Figure 4) showed less heterogeneity on isoelectric focussing (Figure 4 inset) and SDS-PAGE (data not shown), and

produced crystals of sufficient quality for structure determination (see below).

Crystals were grown by the hanging drop vapour diffusion method using purified protein concentrated in Centricon 10 concentrators (Amicon Inc, USA) to 5-10 mg/ml in 10-20 mM Tris-HCl pH 8.0 and 0.02% (w/v) sodium azide, or 100 mM ammonium sulfate and 0.02% (w/v) sodium azide. Crystallization conditions were initially identified using the factorial screen (Jancarik, J. & Kim, S.-H.,1991, J Appl Cryst 24:409-411).and then optimised. Crystals were examined on an M18XHF rotating anode generator (Siemens, Germany) equipped with Franks mirrors (MSC, USA) and RAXIS IIC and IV image plate detectors (Rigaku, Japan).

From the initial crystallization screen of this protein, crystals of about 0.1 mm in size grew in one week. Upon refining conditions, crystals of up to 0.6 x 0.4 x 0.4 mm could be grown from a solution of 1.7-2.0 M ammonium sulfate, 0.1 M HEPES pH 7.5. The crystals varied considerably in shape and diffraction quality, growing predominantly as rhombic prisms with a length to width ratio of up to 5:1, but sometimes as rhombic bipyramids, the latter form being favoured when using material which had been eluted from the Mab 9E10 column at pH 3.0. Each crystal showed a minor imperfection in the form of very faint lines from the centre to the vertices. Protein from dissolved crystals did not appear to be different from the protein stock solution when run on an isoelectric focusing gel. Upon X-ray examination, the crystals diffracted to 3.0-4.0 Å and were found to belong to the space group $P2_12_12_1$ with a = 76.8 Å, b = 99.0 Å, c = 119.6 Å. In the diffraction pattern, the crystal variability noted above was manifest as a large (1-2°) and anisotropic mosaic spread, with concomitant variation in resolution. To improve the quality of the crystals, they were grown in the presence of various additives or were recrystallized. These methods failed to substantially improve the crystal quality although bigger crystals were obtained by recrystallization. The variability in crystal quality appeared to be due to protein heterogeneity, as demonstrated by the observation that more highly purified protein, eluted isocratically from the Resource Q column and showing one major band on isoelectric focusing (Figure 4 inset), produced crystals of sufficient quality for structure determination. These crystals diffracted to 2.6 Å resolution with cell dimensions, a = 77.0 Å, b = 99.5 Å, c = 120.1 Å and mosaic spread of 0.5°. Heavy metal derivatives of the IGF-

1R/462 crystals have been obtained and are leading to the determination of an atomic resolution structure of this fragment, which contains the L1, cysteine-rich and L2 domains of human IGF-1R.

EXAMPLE 2

Structure of the IGF-1R/1-462

Crystals were cryo-cooled to-170°C in a mother liquor containing 20% glycerol, 2.2 M ammonium sulfate and 100 mM Tris at pH 8.0. Native and derivative diffraction data were recorded on Rigaku RAXIS IIc or IV area detectors using copper Kα radiation from a Siemens rotating anode generator with Yale/MSC mirroroptics. The space group was $P2_12_12_1$ with a = 77.39 Å, b = 99.72 Å, and c = 120.29 Å. Data were reduced using DENZO and SCALEPACK (Otwinowski, Z. & Minor, W., 1996, Mode.Meth. Enzym. 276:307-326). Diffraction was notably anisotropic for all crystals examined.

Phasing by multiple isomorphous replacement(MIR) was performed with PROTEIN (Steigeman, W. Dissertation (Technical Univ. Munich, 1974) using anomalous scattering for both UO2 and PIP derivatives. Statistics for data collection and phasing are given in Table 1. In the initial MIR map regions of protein and solvent could clearly be seen, but the path of the polypeptide was by no means obvious. That map was subject to solvent flattening and histogram matching in DM (Cowtan, K.,1994, Joint CCP4 and ESF-EACBM newslett. Protein Crystallogr. 31:34-38). The structure was traced and rebuilt using O (Jones, T. A., et al., 1991, Acta Crystallogr. A47:110-119) and refined with X-PLOR 3.851 (Brunger, A. T., 1996, X-PLOR ReferenceManual 3.851, Yale Univ., New Haven, CT). After 5 rounds of rebuilding and energy minimisation the R-factor dropped to 0.279 and Rfree = 0.359 for data 7-2.6 Å resolution. The current model contains 458 amino acids and 3 N-linked carbohydrates but no solvent molecules. For residues with B(Ca) > 70, Å atomic positions are less reliable (37-42, 155-159, 305, 336-341, 404-406,453-458). There is weak electron density for residues 459-461, but the c-myc tail appears completely disordered.

The 1-462 fragment consists of the N-terminal three domains of IGF-1R (L1, cys-rich, L2), and contains regions of the molecule which dictate ligand specificity (17-23). The molecule adopts a reasonably extended structure (approximately 40 x 48 x 105 Å) with domain 2 (cys-rich region)

making contact along the length of domain 1 (L1) but very little contact with the third domain (L2) (see Figure 5). This leaves a space at the centre of the molecule of approximately 24 Å x 24 Å x 24 Å which is bounded on three sides by the three domains of the molecule. The space is of sufficient size to accommodate the ligand, IGF-1.

Table 1 Summary of Crystallographic data

Data set[a]	Resol. (Å)	Mean I/s	R_{merge}[b]	Completeness (multiplicity)	No. of sites	R_{cullis}[c]	Phasing power[d]	FOMe
Native	2.6	18.7	0.064	0.996 (4.1)				0.47 / 0.71
PIP	3.0	15.8	0.060	0.982 (2.2)	3	0.66	1.71	
UO2Ac2	4.5	7.5	0.095	0.989 (2.3)	2	0.82	1.17	

Refinement resolution (Å)	No of refl. (free)	No. of Atoms	R_{cryst}[f]	R_{free}[f]	Bonds[g] (Å)	Angles[g] (Å)
7.0-2.6	24270 (2693)	3903	0.237	0.304	0.017	0.048

[a] PIP, Di-μ-iodobis(ethylenediamine)diplatinum dinitrate; UO_2Ac_2, Uranyl acetate.

[b] $R_{merge} = \Sigma_h\Sigma_j |I_{h,j}-I_h| / \Sigma_h\Sigma_j I_h$, where $I_{h,j}$ is an intensity measurement j and I_h is the mean intensity for that reflection **h**.

[c] $R_{cullis} = \Sigma_h ||F_{PH}-F_P| - |F_{Hcalc}||/\Sigma_h||F_{PH}|-|F_P||$, where F_{PH}, F_P and F_{Hcalc} are, respectively, derivative, native and heavy atom structure factors for centric reflections **h**.

[d] Phasing power $= \Sigma_h |F_{Hcalc}|/\Sigma_h\varepsilon$, where F_{Hcalc} is defined above and ε is the lack of closure.

[e] FOM(figure of merit) = $<\cos(\Delta\alpha_h)>$, where $\Delta\alpha_h$ is the error in the phase angle for reflection **h**. Values are given before and after density modification at 3.0 and 2.8 Å resolution, respectively.

[f] R_{cryst} and R_{free} are defined in Brunger, A.T. *XPLOR* reference manual 3.851 (Yale Univ., New Haven, CT, 1996)

8 r.m.s. deviation from ideal bond and angle-related (1-3) distances.

The L domains

Each of the L domains (residues 1-150 and300-460) adopts a compact shape (24 x 32 x 37 Å) consisting of a single-stranded right handed β-helix and capped on the ends by short α-helices and disulfide bonds. The body of the domain looks like a loaf of bread, with the base formed from a flat six-stranded β-sheet, 5 residues long and the sides being β-sheets three residues long (Figures 5 & 6). The top is irregular, but in places is similar for the two domains. The two domains are superposable with an rms deviation in Cα positions of 1.6 Å for 109 atoms (Figure 7). Although this fold is reminiscent of other β-helix proteins it is much simpler and smaller with very few elaborations, and thus it represents a new superfamily of domains. One notable difference between the two domains is that the indole ring of Trp 176 from the cys-rich region (Figure 6b) is inserted into the hydrophobic core of L1, and the C-terminal helix is only vestigial (Figure 8). For the insulin receptor family the sequence motif of residues which form the Trp pocket in L1 does not occur in L2 (Figure 6a). However in the EGF receptor, which has an additional cys-rich region after the L2 domain (14, 15), the pocket motif can be found in both L domains and the Trp is conserved in both cys-rich regions (Figure 6b).

The repetitive nature of the β-helix is reflected in the sequence and the first five turns were correctly identified by Bajaj, M., et al. (1987, Biochim.Biophys. Acta 916:220-226), the conserved Gly residues being found in turns making one bottom edge of the domain. However, their conclusions about the fold were incorrect. The"helix-like" repeat is actually a pair of bends at the top edge of the domain. In their Motif V, the Gly is not in a bend but is followed by the insertion of a conserved loop of 7-8 residues (see Figure 6a). Glycine is structurally important in the Gly bends as mutation of these residues compromises folding of the receptor [van der Vorm, E.R., et al., 1992, J. Biol. Chem. 267, 66-71; Wertheimer, E. et al., 1994, J. Biol. Chem. 269, 7587-7592].

Comparison of the L domains with other right-handed β-helix structures such as pectate lyase (Yoder, M. D., et al., 1993,.Structure, 1:241-251-1507) and the p22 tailspike protein (Steinbacher, S., et al., 1997, J.Mol. Biol. 267:865-880) shows some striking similarities as well as differences. In

all cases the ends of the domain are capped by α-helices, but the L domains also have a disulphide bond at each end to hold the termini. The other β-helix domains are considerably longer and have significant twist to their sheets, while the L domains have flat sheets. Although the sizes of the helix repeats are similar (here 24-25 residues vs 22-23 for pectate lyase) the cross-sections are quite different. The L domains have a rectangular cross-section, while pectate lyase and p22 tailspike protein are V-shaped, and have many, and sometimes quite large, insertions (Yoder, M. D., et al., 1993, Structure, 1:241-251-1507; Steinbacher, S., et al., 1997, J.Mol. Biol. 267:865-880). In the hydrophobic core a common feature is the stacking of aliphatic residues from successive turns of the β-helix, and near the C-terminus of each L domain there is also a short Asn ladder, reminiscent of the long Asn ladder observed in pectate lyase (Yoder, M. D., et al., 1993, Structure 1:241-251-1507). On the opposite side of the L domains the Gly bend, as well as the two bends and sheet preceding it, have no counterpart in the other β-helix domains. Thus although the L domains are built on similar principles to the other β-helix domains they constitute a separate superfamily.

The cys-rich domain

The cys-rich domain is composed of eight disulfide-bonded modules (Figure 6b), the first of which sits at the end of L1, while the remainder make a curved rod running diagonally across L1 and reaching to L2 (Figure 5). The strands in modules 2-7 run roughly perpendicular to the axis of the rod in a manner more akin to laminin (Stetefeld, J., et al.,1996, J.Mol.Biol. 257:644-657) than to TNF receptor (Banner, D. W., et al., 1993, Cell, 73:431-445), but the modular arrangement of the cys-rich domain is different to those of other cys-rich proteins for which structures are known. The first 3 modules of IGF-1R have a common core, containing a pair of disulfide bonds, but show considerable variation in the loops (Figure 6b). The connectivity of these modules is the same as in the first half of EGF (Cys 1-3and 2-4), but their structures do not appear to be closely related to any member of the EGF family. Modules 4 to 7 have a different motif, a β-finger, and best match residues 2152-2168 of fibrillin (Dowling, A. K., et al., 1996. Cell, 85:597-605). Each is composed of three polypeptide strands, the first and third being disulfide bonded and the latter two forming a β-ribbon. The β-ribbon of each β- finger module lines up antiparallel to form a tightly twisted 8-stranded β-sheet (Figures 5 and 8). Module 6 deviates from the common pattern,with

the first segment being replaced by an α-helix followed by a large loop that is likely to have a role in ligand binding (see below). As module 5 is most similar to module 7 it is possible that the four modules arose from serial gene duplications. The final module is a disulfide-linked bend of five residues.

The fact that the two major types of cys-rich modules occur separately implies that these are the minimal building blocks of cys-rich domains found in many proteins. Although it can be as short as 16 residues, the motif of modules 4-7 is clearly distinct, and capable of forming a regular extended structure. Thus cys-rich domains such as these can be considered as being made of repeat units each composed of a small number of modules.

Hormone binding

Attempts have been made to locate the IGF-1 (and insulin) binding site by examining natural (Taylor, S. I., 1992, Diabetes, 41:1473-1490) and site-directed mutants (Williams, P. F., et al., 1995, J. Biol. Chem. 270:3012-3016; Mynarcik, D. C et al., 1996, J. Biol. Chem. 271:2439-2442; Mynarcik, D. C., et al., 1997, J. Biol. Chem. 272:2077-2081), chimeric receptors (Andersen, A. S., et al., 1990, Biochemistry 29:7363-7366; Gustafson, T. A., & Rutter, W. J., 1990, J. Biol. Chem. 265:18663-18667; Schäffer, L., et al.,1993, J. Biol. Chem. 268:3044-3047; Schumacher, R., 1993, J. Biol. Chem. 268:1087-1094; Kjeldsen, T., et al., 1991, Proc. Natl Acad. Sci. USA, 88:4404-4408) and by crosslinking studies (Wedekind, F., et al., 1989, Biol. Chem Hoppe-Seyler, 370:251-258; Fabry, M., 1992, J. Biol. Chem. 267:8950-8956; Waugh, S. M., et al., 1989, Biochemistry, 28:3448-3458; Kurose, T., et al., 1994),.J. Biol. Chem.269:29190-29197-34). IGF-1R/IR chimeras not only show which regions of the receptors account for ligand specificity, but also provide an efficient means of identifying some parts of the hormone binding site. Paradoxically, regions controlling specificity are not the same for insulin and IGF-1. Replacing the first 68 residues of IGF-1R with those of IR confers insulin-binding ability on the chimeric IGF-1R (Kjeldsen, T., et al., 1991, Proc. Natl Acad. Sci. USA, 88:4404-4408), and replacing residues 198-300 in the cys-rich region of IR with the corresponding residues 191-290 of IGF-1R allows the chimeric receptor to bind IGF-1 (Schäffer, L., et al.,1993, J. Biol. Chem. 268:3044-3047). Thus a receptor can be constructed which binds both IGF-1 and insulin with near native affinity. From the structure it is clear that if the hormone bound in the central space it could contact both these regions.

From analysis of a series of chimeras examined by Gustafson, T. A., & Rutter, W. J. (J. Biol. Chem. 265:18663-18667, 1990), the specificity determinant in the cys-rich region can be limited further to residues 223-274. This region corresponds to modules 4-6, and includes a large and somewhat mobile loop (residues 255-263, mean B[Cα atoms] = 57 Å2) which extends into the central space (see Figure 5). In IR this loop is four residues bigger, and is stabilised by an additional disulfide bond (Schäffer, L. & Hansen, P.H.,1996, Exp. Clin. Endocrinol. Diabetes, 104: Suppl. 2, 89). The larger loop of IR may serve to exclude IGF-1 from the hormone binding site while allowing the smaller insulin molecule to bind. It is interesting to note that mosquito IR homologue, which has a loop two residues larger than the mammalian IRs, also appears to bind insulin but not IGF-1 (Graf, R., et al., 1997, Insect Molec.Biol. 6:151-163). Analysis of the structure indicates that the insulin/IGF-1 specificity is controlled by residues in this loop (amino acids 253-272 in IGF-1R; amino acids 260-283 in IR)

As chimeras only address residues which differ between the two receptors, a more precise analysis of the site can be obtained from single site mutants. In particular, from an alanine-replacement study, four regions of L1 important for insulin binding were identified (Williams, P. F., et al., 1995, J. Biol. Chem. 270:3012-3016). The first three are at similar positions on successive turns of the β-helix and the fourth lies on the conserved bulge on the large β-sheet. Thus there is a footprint for insulin binding to the L1 domain which lies on the first half of the large β-sheet facing into the central space. Residues further along the sheet which are conserved in IGF-1R could also be important. The conservative substitution of leucine for methionine at residue 119 of IR (113 of IGF-1R) causes a mild form of leprechaunism [Hone, J. et al., 1994, J. Med. Genet. 31, 715-716]. This residue is buried, and the mutation could perturb neighbouring residues to affect insulin binding.

The axis of the L2 domain is perpendicular to that of the L1 domain, and the N-terminal end of its β-helix is presented to the hormone-binding site. On this face of the L2 domain the only mutation studied so far is the naturally occurring IR mutant, S323L, which gives rise to Rabson-Mendehall syndrome and severe insulin resistance (Roach, P.,1994, Diabetes 43:1096-1102). As this mutant only affects insulin binding and not cell-surface expression, residue 323 of IR (residue 313 of IGF-1R) is probably at or near the binding site. Structurally this residue lies in the middle of a region

(residues 309-318 of IGF-1R) which is conserved in both IR and IGF-1R, and the surrounding region, 332-345 (of IGF-1R), is also quite well conserved in the these receptors (Figure 6a). Therefore this region is quite likely to form part of the hormone-binding site, but would not have been detected by chimeras. It is interesting to note that in this region IRR is not as well conserved as the other two receptors (Shier, P. & Watt, V.M., 1989, J.Biol.Chem. 264:4605-14608).

The distance from this putative hormone-binding region on L2 to that found on L1 is about 30 Å (Figure 5). Thus L1 and L2 appear too far apart to bind IGF-1 or insulin. However, in the crystal structure there is a deep cleft between part of the cys-rich domain (residue 262)and L2 (residue 305), and this cleft is occupied by a loop from a neighbouring molecule. Thus it seems probable that the position of the L2 domain in the receptor structure or the hormone-receptor complex adopts a different position with respect to the cys-rich domain than that found in the crystal. The movement required to bring L2 sufficiently close to L1 is small, namely a rotation of approximately 25° about residue 298.

A number of IR mutants have been identified which constitutively activate the receptor, and the majority of these are found in the α chain. Curiously all α chain mutants involve changes to or from proline or the deletion of an amino acid, implying that they cause local structural rearrangements. The mutation R86N is similar to wild type, but R86P reduces cell-surface expression and insulin binding while constitutively activating autophosphorylation [Grønskov, K. et al., 1993, Biochem. Biophys. Res. Commun. 192, 905-911]. The proline mutation probably disturbs residues preceding 87 which lie in the interface between the L1 and cys-rich domains, but it could also affect insulin binding. In the cys-rich domain residues 233, 281, 244 and 247 of IR are not conserved in IGF-1R (Figure 6b), yet L233P [Klinkhamer, M.P. et al., 1989, EMBO J. 8, 2503-2507], deletion of N281 [Debois-Mouthon, C. et al., 1996, J. Clin. Endochronol. Metab. 81, 719-727] or the triple mutant P243R, P244R and H247D [Rafaeloff, R. et al., 1989, J. Biol. Chem. 264, 15900-15904] cause constitutive kinase activation. Due to their locations each of these three mutants appears likely to compromise the folding of a β-finger domain and, in turn, the structural integrity of the rod-like cys-rich domain. The structural ramifications of these mutations could be significant for the whole receptor ectodomain, as disturbing the L1/cys-

rich interface or distorting the rod-like domain could affect the relative position of L1 and the cys-rich domain in this context.

L1 has been further implicated, as deletion of K121 on the opposite side of L1 from the cys-rich domain was also found to cause autophosphorylation [Jospe, N. et al., 1994, J. Clin. Endochronol. Metab. 79, 1294-1302]. By contrast this mutation does not affect insulin binding. Thus a possible mechanism emerges for insulin binding and signal transduction. When insulin binds between L1 and L2 it modifies the relative position of L1 and the cys-rich domain in the receptor, perhaps by hinge motion between L2 and the cys-rich domain like that suggested above, and the structural rearrangement is transmitted across the plasma membrane. In the absence of insulin the same signal can be initiated by mutations in the cys-rich region or at the L1/cys-rich interface, but at the expense on insulin binding. The signal can also be initiated more directly by mutations on the opposite side of L1 which affect the interaction of L1 with other parts of the ectodomain, possibly the other half of the receptor dimer.

Ligand Studies

Although there is no structural information about an IGF-1/IGF-1R complex a number of studies have probed the nature of this interaction. Results from cross-linking experiments with IGF-1 and insulin and their cognate receptors are consistent with the hormone binding site proposed above. For example B29 of insulin can be cross-linked to the cys-rich region (residues 205-316((Yip, C. C., et al., 1988, Biochim. Biophys. Res. Commun. 157:321-329) or the L1 domain (Wedekind, F., et al., 1989, Biol. Chem Hoppe-Seyler, 370:251-258). However, these two regions are reasonably well separated, and those studies may indicate that B29 is mobile. Other studies unfortunately do not map the site any more precisely.

Analogues and site-directed mutants of IGF-1 and IGF-2 have been more fruitful. IGF-1 and IGF-2 contain two extra regions relative to insulin, the C region between B and A and a D peptide at the C-terminus. For IGF-1, replacement of the C region by a four Gly linker reduced affinity for IGF-1R by a factor of 40 but increased affinity for IR 5-fold (Bayne, M.L.,et al., 1988, J. Biol.Chem. 264:11004-11008). Changes in affinity are consistent with the deletion in IGF-1 complementing differences in the cys-rich regions of IGF-1R and IR noted above. Mutation of residues either side of the C region (residue 24 for IGF-1 [Cascieri, M.A., et al., 1988, Biochemistry 27:3229-

3233], residues 27,43 for IGF-2, [Sakano, K., et al., 1991, J. Biol. Chem. 266:20626-20635]) also has deleterious effects on the affinity of the hormone for IGF-1R. as has truncation of the nearby D peptide in IGF-2 (Roth, B.V., et al., 1991, Biochem. Biophys. Res. Commun. 181:907-914).

Insulin has been extensively mutated. Binding studies [summarised in Kristensen, C. et al., 1997, J. Biol. Chem. 272, 12978-12983] indicate that insulin may bind its receptor via a hydrophobic patch (residues A2, A3, A19, B8, B11, B12, B15 and possibly B23 & B24). However this patch is normally buried, and requires the removal of the B chain's C-terminus from the observed position. Assuming IGF-1, IGF-2 and insulin bind their receptors in the same orientation, these data suggest an approximate orientation for the hormone when bound to the receptor.

One notable feature of IGF-1 and IGF-2 is the large number of charged residues and their uneven distribution over the surface. Basic residues are predominantly found in the C region and, in solution, this region is not well ordered in either IGF-1 or -2 (Sato, A., et al., 1993, Int J Peptide Protein Res. 41:433-440; Torres, A. M., et al., 1995,J. Mol. Biol. 248:385-401). In contrast the binding site of the receptor has a sizable patch of acidic residues in the corner where the cys-rich domain departs from L1. Other acidic residues which are specific to this receptor are found along the inside face of the cys-rich domain and the loop (residues 255-263) extending from module 6. Thus it is possible that electrostatic interactions play an important part in IGF-1 binding, with the C region binding to the acidic patch of the cys-rich region near L1 and the acidic patch on the other side of the hormone directed towards a small patch of basic residues (residues 307-310) on the N-terminal end of L2.

Although the structure of this fragment gives significant information about the nature of the hormone binding site, residues outside this region have also been shown to affect binding of ligand. A number of studies have implicated residues 704-715 of IR (Mynarcik, D. C et al., 1996, J. Biol. Chem. 271, 2439-2442; Kurose, T., et al., 1994, J. Biol. Chem.269:29190-29197). These residues could contact insulin on one of the sides left open in the current structure. Using insulin labelled at the B1 residue, Fabry, M., et al.,(1992, J. Biol. Chem. 267:8950-8956) cross-linked insulin to the fragment 390-488, part of which is not near the site as described. The explanation for this could be either the region 390-488 reaches back to the hormone binding

site, or this region could contact another hormone bound to the other half of the receptor. Further structural information is needed to establish how these other regions contact the hormone and to elucidate how binding of the hormone is communicated to the kinase inside the cell.

The structure of the L1-cys-rich-L2 domains of IGF-1R presented here represents the first structural information for the extracellular portion of a member of the insulin receptor family. The L domains display a novel fold which is common to the EGF receptor family, and the modular architecture of the cys-rich domain implies that smaller building blocks should be used to describe the composition of cysteine-rich domains. This fragment contains the major specificity determinants of receptors of this class for their ligands. It has an elongated structure with a space in the middle which could accommodate the ligand. The three sides of this site correspond to regions which have been implicated in hormone binding. Although other sites are present in the receptor ectodomain which interact with the ligand, this structure gives us an initial view of how the insulin, IGF-1 and IGF-2 might interact with their cell surface receptors to control their metabolic and mitogenic effects

Such information will provide valuable insight into the structure of the corresponding domains of the IR and insulin receptor-related receptor as well as members of the related EGFR family (Bajaj, M., et al., 1987, Biochim Biophys Acta 916:220-226; Ward, C. W. et al., 1995, Proteins: Struct Funct Genet 22:141-153).

EXAMPLE 3

Prediction of 3D Structure of the Corresponding Domains of IRR and IR Based on Structure of IGF-1R Fragment.

The sequence identities between the different members of the insulin receptor family are sufficient to allow accurate sequence alignments to facilitate 3D structure predictions by homology modelling. The alignments of the ectodomains of human IGF-1R, IR, and IRR are shown in Figure 9.

EXAMPLE 4

Single-Molecule Imaging of Human Insulin Receptor Ectodomain and its Fab Complexes

Cloning and expression of hIR -11 ectodomain protein

A full length clone of the human IR exon -11 form (hIR -11) was prepared by exchanging an Aat II fragment, nucleotides 1195 to 2987 , of the

exon +11 clone (plasmid pET; Ellis et al., 1986; gift from Dr W. J. Rutter, UCSF) of hIR (Ebina et al., 1985, *Cell* **40**, 747-758) with the equivalent Aat II fragment from a plasmid (pHIR/P12-1, ATCC 57493) encoding part of the extracellular domain and the entire cytoplasmic domain of hIR -11 (Ullrich et al., 1985, *Nature* **313** , 756-761). The ectodomain fragment of hIR -11 (2901 bp, coding for the 27 residue signal sequence and residues His1-Asn914) was produced by SalI and SspI digestion and inserted into the mammalian expression vector pEE6.HCMV-GS (Celltech Limited, Slough, Berkshire, UK) into which a stop codon linker had been inserted, as described previously (Cosgrove et al., 1995, *Protein Expression and Purification* **6**, 789-798) for the hIR exon +11 ectodomain.

The resulting recombinant plasmid pHIR II (2 μg) was transfected into glycosylation-deficient Chinese hamster ovary (Lec 8) cells (Stanley, 1989, *Molec. Cellul. Biol.* **9**, 377-383) with Lipofectin (Gibco-BRL). After transfection, the cells were maintained in glutamine-free medium GMEM (ICN Biomedicals, Australia) as described previously (Bebbington & Hentschel, 1987, In *DNA Cloning* (Glover, D., ectodomain.), Vol III, Academic Press, san Diego; Cosgrove et al., 1995, *Protein Expression and Purification* **6**, 789-798). Expressing cell lines were selected for growth in GMEM with 25 μM methionine sulphoximine (MSX, Sigma). Transfectants were screened for protein expression using sandwich ELISA with anti-IR monoclonal antibodies 83-7 and 83-14. Metabolic labelling of cells, immunoprecipitations, insulin binding assays and Scatchard analyses were performed as described previously for the exon +11 form of hIR ectodomain (Cosgrove et al., 1995, , *Protein Expression and Purification* **6**, 789-798).

hIR -11 ectodomain production and purification

The selected clone (inoculum of 1.28 x 108 cells) was grown in a spinner flask packed with 10 g of Fibra-cel disc carriers (Sterilin, U.K.) in 500 ml of GMEM medium containing 10% fetal calf serum (FCS) and 25 μM MSX. Selection pressure was maintained for the duration of the culture.

Ectodomain was recovered from harvested medium by affinity chromatography on immobilized insulin, and further purified by gel filtration chromatography on Superdex S200 (Pharmacia; 1 x 40 cm) in Tris-buffered saline containing 0.02% sodium azide (TBSA) as described previously (Cosgrove et al., 1995, *Protein Expression and Purification* **6**, 789-798). Solutions of purified hIR -11 ectodomain were stored at 4° C prior to use.

Production of Fab fragments and their complexes with ectodomain

Purification of Mabs 83-7, 83-14 and 18-44 from ascites fluid by affinity chromatography using Protein A-Sepharose, and the production of Fabs, were based on the methodologies described in Coligan et al.,1993, Current Protocols in Immunology, Vol 1, pp 2.7.1-2.8.9, Greene Publishing Associates & Wiley - Interscience, John Wiley and Sons. Fab was produced from monoclonal antibody by mercuripapain digestion for 1-4 h, followed by gel filtration on Superdex S200. Products were monitored by reducing and non-reducing SDS-PAGE. For 83-7 Mab, an IgG Type 1 monoclonal antibody, the bivalent (Fab)2' isolated by this method was reduced to monovalent Fab 83-7 by mild reduction with mM L-cysteine.HCl in 100 mM Tris pH 8.0 (Coligan et al., 1993, Current Protocols in Immunology, Vol 1, pp 2.7.1-2.8.9, Greene Publishing Associates & Wiley - Interscience, John Wiley and Sons).

Complexes of Fab with hIR -11 ectodomain were produced by mixing ~ 2.5 to 3.5 molar excess of Fab with hIR -11 ectodomain at ambient temperature in TBSA at pH 8.0. After 1-3 h, the complex was separated from unbound Fab by gel filtration over a Superdex S200 column in the same buffer.

Electron microscopy

Uncomplexed hIR -11 ectodomain and the Fab complexes described above were diluted in phosphate-buffered saline (PBS) to concentrations of the order of 0.01-0.03 mg/ml. Prior to dilution, 10% glutaraldehyde (Fluka) was added to the PBS to achieve a final concentration of 1% glutaraldehyde. Droplets of ~ 3ml of this solution were applied to thin carbon film on 700-mesh gold grids after glow-discharging in nitrogen for 30 s. After 1 min. the excess protein solution was drawn off and followed by application and withdrawal of 4-5 droplets of negative stain [2% uranyl acetate (Agar), 2% uranyl formate (K and K), 2% potassium phosphotungstate (Probing and Structure) adjusted to pH 6.0 with KOH, or 2% methylamine tungstate (Agar) adjusted to pH 6.8 with NH4OH]. In the case of both uranyl acetate and uranyl formate staining, an intermediate wash with 2 or 3 droplets of PBS was included prior to application of the stain. The grids were air-dried and

then examined at 60kV accelerating voltage in a JEOL 100B transmission electron microscope at a magnification of 100,000x. It was found that there was a typical thickness of negative stain in which Fabs were most easily seen. Hence areas for photography had to be chosen from particular zones of the grid. Electron micrographs were recorded on Kodak SO-163 film and developed in undiluted Kodak D19 developer. The electron-optical magnification was calibrated under identical imaging conditions by recording single-molecule images of the antigen-antibody complex of influenza virus neuraminidase heads and NC10 MFab (Tulloch et al., 1986, *J.Mol. Biol.* **190**, 215-225; Malby et al., 1994, *Structure*, **2**, 733-746).

Image processing

Electron micrographs showing particles in a limited number of identifiable projections were chosen for digitisation. Micrographs were digitised on a Perkin-Elmer model 1010 GMS PDS flatbed scanning microdensitometer with a scanning aperture (square) size of 20 mm and stepping increment of 20 mm corresponding to a distance of 0.2 nm on the specimen. Particles were selected from the digitised micrograph using the interactive windowing facility of the SPIDER image processing system (Frank et al., 1996, *J. Struct. Biol.* **116**, 190-199). Particles were scaled to an optical density range of 0.0 - 2.0 and aligned by the PSPC reference-free alignment algorithm (Marco et al., 1996, *Ultramicroscopy*, **66**, 5-10). Averages were then calculated over a subset of correctly aligned particles chosen interactively as being representative of a single view of the particle. The final average image presented here is derived from a library of 94 images.

Biochemical characterization of expressed hIR -11 ectodomain

The recombinant protein examined corresponded to the the first 914 residues of the 917 residue ectodomain of the exon -11 form of the human insulin receptor (Ullrich et al., 1986, *Nature* **313** , 756-761). Expressed protein was shown, by SDS-PAGE and autoradiography of immunoprecipitated product from metabolically labelled cells, to exist as a homodimeric complex of ~270 - 320 kDa apparent mass, which dissociated under reducing conditions into monomeric α and β' subunits of respective apparent mass ~120 kDa and ~35 kDa (data not shown).

Purified hIR -11 ectodomain, expressed in Lec8 cells and purified by affinity chromatography on an insulin affinity column, eluted as a symmetrical peak on a Superdex S200 gel filtration column (Figure 10). The

protein eluted with an apparent mass of ~400 kDa, calculated from a standard curve generated by the elution positions of standard proteins (not shown). As expected for protein expressed in Lec 8 cells, whose glycosylation defect produces truncated oligosaccharides (Stanley, 1989, . *Molec. Cellul. Biol.* **9**, 377-383), this value is less than the apparent mass (450 - 500 kDa) reported for hIR +11 ectodomain expressed in wild-type CHO-K1 cells (Johnson et al., 1988, *Proc. Natl Acad. Sci USA* **85**, 7516-7520; Cosgrove et al., 1995, *Protein Expression and Purification* **6**, 789-798).

Radioassay of insulin binding to purified ectodomain gave linear Scatchard plots and Kd values of 1.5 - 1.8 x 10-9 M, similar to the values of 2.4 - 5.0 x 10-9 M reported for the hIR -11 ectodomain (Andersen et al., 1990, *Biochemistry* **29**, 7363-7366; Markussen et al., 1991, *J. Biol. Chem.* **266**, 18814-18818; Schaffer, 1994, *Eur. J. Biochem.* **221**, 1127-1132) and the values of ~1.0 - 5.0 x 10-9 M reported for the hIR +11 ectodomain (Schaefer et al., 1992, *J. Biol. Chem.* **267**, 23393-23402; Whittaker et al., 1994, *Molec. Endocrinol.* **8**, 1521-1527; Cosgrove et al., 1995, *Protein Expression and Purification* **6**, 789-798).

Expression of hIGF-1R ectodomain

Cloning, expression and purification of this protein used elements common to those described for hIR -11 ectodomain (Cosgrove et al., 1995, *Protein Expression and Purification* **6**, 789-798), and resulted in purified product that was recognised by receptor-specific Mabs 17-69, 24-31 and 24-60 (Soos et al., 1992, *J. Biol. Chem.* **267**, 12955-63) and was composed of α and β' subunits of mass similar to those of hIR ectodomain.

Preparation of hIR -11 ectodomain/MFab complexes

A complex of hIR -11 ectodomain and Fab from antibody 83-14 eluted as a symmetrical peak of 460 -500 kDa (Figure 10), as did complexes generated from a mixture of hIR -11 ectodomain with Fab from antibody 18-44 and a mixture of hIR -11 ectodomain with Fab 83-7 (not shown). A co-complex of ectodomain with Fabs from antibodies 18-44 and 83-14 eluted at ~ 620 kDa, as did a co-complex with MFabs 83-14/83-7 and another with MFabs 83-7/18-44 (not shown). A complex of hIR -11 ectodomain with all three MFab derivatives, 18-44, 83-7 and 83-14, eluted at an apparent mass of ~ 710 kDa (Figure 10).

Electron microscopy

Imaging of hIR -11 and hIGF-1R ectodomains

Single-molecule imaging of uncomplexed dimeric hIR -11 ectodomain was carried out under a variety of negative staining conditions, which emphasised different aspects of the structure of the molecular envelope. Images obtained by this investigation are depicted in Figure 11.

The least aggressive or penetrative stain was potassium phosphotungstate (KPT) , which revealed consistent globular particles with very little internal structure other than a suggestion of a division into two parallel bars. Staining with methylamine tungstate also revealed the parallel bar images.

Further investigation using progressively more penetrative, but also potentially more disruptive, stains confirmed the observations above. Staining with uranyl acetate and uranyl formate showed the separation of the parallel bars most clearly, but uranyl acetate showed evidence of disrupting the structure of the particles, i.e. a decrease in the consistency of the particle shape and a tendency for particles to look unravelled or denatured despite having been subjected to chemical cross-linking prior to staining. In areas of thicker stain, parallel bars predominated, whereas in more thinly stained regions, U-shaped particles could be identified, sometimes outnumbering the parallel-bar structures (see Figure 11).

Imaging of hIR -11 ectodomain complexed with 83-7 MFab

This complex was particularly noteworthy for the consistency of the form of the particles, especially under the gentler staining conditions afforded by stains such as KPT and methylamine tungstate. The particles were interpreted as having been restricted in the views they presented, after air-drying on the carbon support film, by the almost diametrically opposite binding of the two Fab arms to the antigen to form a highly elongated complex structure. Under these conditions three distinct views could be recognised (see Figure 11). Two views (interpreted as top-down/bottom-up) show the Fab arms displaced clockwise or anti-clockwise as extensions of the parallel plates with two-fold symmetry. The third view shows an image with the two Fab arms in line roughly through the centre of the receptor on its opposite sides, interpreted as a side projection of binding half-way up the plates.

The use of aggressive uranyl stains operating at lower pHs revealed internal structure of the molecular envelope at the expense of consistency of the particle morphology. For example, staining with uranyl acetate or uranyl

formate showed that parallel bars can be seen in particles in which the Fab arms are displaced either clockwise or anticlockwise but not where the intermediate central or axial position of the two Fab arms is presented in projection. These observations show 83-7 MFab binding roughly half-way up the side-edge of each hIR -11 ectodomain plate. The epitope recognised by Mab 83-7 has been mapped to the cys-rich region, residues 191-297, by analysis of chimeric receptors (Zhang and Roth, 1991, *Proc. Natl. Acad. Sci. USA* **88**, 9858-9862).

Imaging of hIR -11 ectodomain complexed with either 83-14 MFab or 18-44 MFab

Complexes were formed with Fabs from the most insulin-mimetic antibody Mab 83-14. Projections showing the Fab arms bound to and extending out from near the base of the U-shaped particles were identified. A second field of particles showed objects composed of two parallel bars as observed for the undecorated ectodomain, with Fab arms projecting obliquely from diametrically opposite extremities (see Figure 11). Similar but less definitive images were also seen when MFab 18-44 was bound to hIR -11 ectodomain. The epitope for Mab 83-14 is between residues 469-592 (Prigent et al., 1990) in the connecting domain. This domain contains one of the disulphide bonds (Cys524-Cys524) between the two monomers in the IR dimer (Schaffer and Ljungqvist, 1992, *Biochem. Biophys. Res. Commun.* **189**, 650-653). The epitope for Mab 18-44 is a linear epitope, residues 765-770 (Prigent et al., 1990, . *J. Biol. Chem.* **265**, 9970-9977) in the β-chain, near the end of the insert domain (O'Bryan et al., 1991, *Mol. Cell. Biol.* **11**, 5016-5031). The insert domain contains the second disulphide bond connecting the two monomers in the IR dimer (Sparrow et al., 1997, *J. Biol. Chem.*, 272, 29460-29467).

Imaging of hIR -11 ectodomain co-complexed with two different MFabs per monomer

The double complex of hIR -11 ectodomain with MFabs 83-7 and 18-44 was stained with 2% KPT at pH 6.0, and revealed the molecular envelopes. The particle appears complex in shape, and can assume a number of different orientations on the carbon support film, giving rise to a number of different projections in the micrograph. The predominant view is of an asymmetric X-shape (some examples circled). It shows the 83-7 MFab arms bound at opposite ends of the parallel bars with the two 18-44 MFabs

appearing as shorter projections extending out from either side of each ectodomain.

Images of the double complex of hIR -11 ectodomain with 83-7 and 83-14 MFabs gave X-shaped images similar to those seen with the 83-7/18-44 double complex. In contrast the double complex of hIR -11 ectodomain with 18-44 and 83-14 MFabs did not present the characteristic asymmetric X-shapes described above. Instead, the molecular envelope appeared to be elongated in many views, with only an occasional X-shaped projection. While a detailed interpretation of these images would be premature, it is clear that MFabs 18-44 and 83-14, two of the more potent insulin mimetic antibodies (Prigent et al., 1990, *J. Biol. Chem.* **265**, 9970-9977), can bind simultaneously to the receptor.

Imaging of hIR -11 ectodomain co-complexed with three different MFabs per monomer

A field of particles from a micrograph of hIR -11 ectodomain were complexed simultaneously with MFabs 83-7, 83-14 and 18-44. In the thicker stain regions the molecular envelope was X-shaped, and looked very similar to that of the double complexes of hIR -11 ectodomain with either 83-7 and 18-44 or 83-7 and 83-14. However, in the more thinly stained regions particles of greater complexity were visible, and it was possible occasionally to identify that there are in fact more than four MFabs bound to the ectodomain dimer.

The single-molecule imaging of hIR -11 ectodomain presented here suggests a molecular envelope for this dimeric species significantly different from that of any previously published study. However, an unequivocal determination of the molecular envelope even from the present study is not entirely straightforward. A major complicating factor here has been the relative fragility of the expressed ectodomain when exposed to the rigors of electron microscope preparation by negative staining. For example, staining with potassium phosphotungstate (KPT, pH 6.0-7.0) frequently suggested a denaturation of the dimeric molecules, but when appropriate conditions were satisfied, good seemingly interpretable molecular envelope images were achieved; staining with methylamine tungstate (pH ~7.0) supported the best KPT molecular envelope images, but had the suggestion of a swelling of the molecular structure at neutral pH; and the acid-pH stains of uranyl acetate (pH ~4.2) and uranyl formate (pH~3.0), with their ability to penetrate the

ectodomain structure, appeared to illuminate not so much the molecular envelope as the zones of high projected protein density within the dimer.

An amalgam of impressions from these various staining regimens has led to the following interpretation of single-molecule images of these undecorated, or naked, dimers: the predominant dimeric molecular image encountered here has been that of "parallel bars"of projected protein density. This view is so predominant, indeed, that it suggests there is either a single preferred orientation of the molecules on the glow-discharged carbon support film, or that this impression of parallel bars of density may represent a mixture of superficially similar structure projections, with the subtleties of these different projections being masked by the relatively coarse resolution of this single-molecule direct imaging. The impression of parallel bars of projected protein density is particularly predominant in regions of thicker negative stain. A second view of the molecular envelope, appreciably less well represented in regions of thicker stain but predominant in regions of thin staining, is that of 'open' U's, or V's. These two views of hIR -11 ectodomain were supported by the single-molecule imaging of hIGF-1R ectodomain under comparable conditions of negative staining.

If the assumption is made that these two recognisable projected views, that of parallel bars and of open U's/V's, are different views of the same dimeric molecule, an assumption strongly supported by the MFab complex imaging, a coarse model of the molecular envelope can be rationalized. The model structure is roughly that of a cube, composed of two almost-parallel plates of high protein density, separated by a deep cleft of low protein main-chain and side-chain density able to be penetrated by stain, and connected by intermediate stain-excluding density near what is assumed here to be their base (that is, nearest the membrane-anchoring region). The width of the low-density cleft appears to be of the order of 30-35Å, sufficient to accommodate the binding of the insulin molecule of diameter ca. 30Å, although we have no electron microscopical evidence to support insulin-binding in this cleft at this stage.

It has been established through imaging of bound 83-7 MFab that there is a dimeric two-fold axis normal to the membrane surface between these plates of density. Occasionally, dimer images display a relative displacement of the bars of density, interpreted here as a limited capacity for a shearing of the interconnecting zone between the two plates along their

horizontal axis parallel to the membrane; other images show bars skewed from parallel, implying a limited capacity for the plates to rotate independently around the two-fold axis, again via this interconnecting zone. These two observations each suggest a relatively flexible connectivity between the dimer plates in the membrane-proximal region of intermediate protein density, which could possibly contribute to the transmembrane signalling process.

The approximate overall measured dimensions of the ectodomain dimer are 110 x 90 x 120Å, calibrated against the dimensions of imaged influenza neuraminidase heads, known from the solved X-ray structure (Varghese et al., 1983, *Nature* **303**, 35-40). It can be noted that there is a compatibility here between the molecular weights and molecular dimensions of these two molecular species: the compact tetrameric influenza neuraminidase heads of Mr ~200 kDa occupy a volume almost 100 x 100 x 60 Å; the more open dimeric insulin receptor ectodomains of similar Mr ~240 kDa imaged here occupy a volume approximately 110 x 90 x 120 Å , roughly twice that of the neuraminidase heads, accommodating the slightly higher molecular weight and substantial central low-density cleft.

The low-resolution roughly cubic compact structure proposed here differs substantially from the T-shaped model proposed by Christiansen et al. (1991, *Proc. Natl. Acad. Sci. U. S. A.* **88**, 249-252) and Tranum-Jensen et al., (1994, *J. Membrane Biol.* **140**, 215-223) for the whole receptor and the elongated model proposed by Schaefer et al. (1992, *J. Biol. Chem.* **267**, 23393-23402) for soluble ectodomain. Significantly, those previous studies did not provide any convincing independent electron microscopical evidence that their imaged objects were in fact insulin receptor.

In the present study, the identity of the imaged molecules as hIR -11 ectodomain has been confirmed by imaging complexes of the dimer with Fabs of the three well-established conformational Mabs against native hIR, 83-7, 83-14 and 18-44 (Soos et al.,1986, *Biochem. J.* **235**, 199-208; 1989, *Proc. Natl Acad. Sci. USA* **86**, 5217-5221), bound singly and in combination. In all these instances, virtually every particle in the field of view exhibited MFab decoration through binding to conformational epitopes, establishing not only the identity of the imaged particles but also the conformational integrity of the expressed ectodomains. Furthermore, the cleanliness and uniformity of these hIR -11 ectodomain preparations, both naked and decorated, visualised

here by electron microscopy demonstrate their high suitability for X-ray crystallization trials.

The known flexibility of the Fab arms exacerbates image-to-image variability beyond the limited extent already described for the undecorated dimeric ectodomains, complicating any precise interpretation of these antigen-antibody complexes. Such molecular flexibility also renders largely impractical any single-molecule computer image averaging to facilitate image interpretation, progressively more so with the higher order antigen-antibody complexes studied here.

The most readily interpretable of these images, showing least image-to-image variability, are those of 83-7 MFab bound to dimers where, fortuitously, the antigen-antibody complex is constrained in its degrees of rotational freedom on the carbon support film. Many projected images show the two Fab arms in line roughly through the centre of the antigen on its opposite sides, interpreted as a side projection of binding half-way up the plates from their membrane-proximal base. Other sub-sets of images show the two Fab arms still parallel but displaced clockwise or anticlockwise with 2-fold symmetry, each Fab approximating an extension of one of the parallel bars of antigen density, interpreted here as representing top or bottom projections along the 2-fold axis. The third projection, along the axis of the Fab arms, could not be sampled here because of the constraining geometry of this molecular complex. These observations suggest binding of 83-7 MFab roughly half-way up the side-edge of the hIR -11 ectodomain plate. This then allows an initial attempt at spatially mapping the 83-7 MFab epitope, which has been sequence-mapped to residues 191-297 in the cys-rich region of the insulin receptor (Zhang and Roth, 1991, *Proc. Natl. Acad. Sci. USA* **88**, 9858-9862). The spatial separation and relative orientations of the two binding epitopes of Mab 83-7 on the hIR -11 ectodomain dimer as indicated here appear inconsistent with the proposal that Mab 83-7 could bind intramolecularly to hIR (O'Brien et al., 1987, *Biochem J.* **6**, 4003-4010).

Decoration of the ectodomain dimer with 83-7 MFab established that the two plates of high protein-density are arranged with 2-fold symmetry. Decoration with either 83-14 or 18-44 MFab , on the other hand, allowed sampling of the third projection of the ectodomain dimer precluded by 83-7 MFab binding. Significantly, this third view established unequivocally the U-shaped projection of the hIR -11 ectodomain dimer, something which was

only able to be assumed with the undecorated ectodomain images. Further, this projection has allowed a rough spatial mapping close to the base of the U-shaped dimer for the epitopes recognised by 83-14 MFab (residues 469-592, connecting domain) and 18-44 MFab (residues 765-770, b-chain insert domain; exon 11 plus numbering, Prigent et al., 1990, *J. Biol. Chem.* **265**, 9970-9977).

Inherent in the model structure is the implication that, with the two-fold axis aligned normal to the membrane surface, the mouth of the low-density cleft where insulin binding may occur would lie most distant from the transmembrane anchor, whilst the zone of intermediate density connecting the two high-density plates would be in close proximity to the membrane. It follows, in this model, that the L1/cys-rich/L2 domains(Bajaj et al., 1997, *Biochim. Biophys. Acta* **916**, 220-226; Ward et al.,1995, *Proteins: Struct., Funct., Genet.* **22**, 141-153), which comprise much of the insulin-binding region (see Mynarcik et al., 1997, . *J. Biol. Chem.* **272**, 2077-2081), most probably lie in the membrane-distal upper halves of the two plates, whilst the membrane-proximal lower halves contain the connecting domains, the fibronectin-type domains, the insert domains and the interchain disulphide bonds (Schaffer and Ljungqvist, 1992, *Biochem. Biophys. Res. Commun.* **189**, 650-653; Sparrow et al., 1997, *J. Biol. Chem.*, 272, 29460-29467). Such a disposition of domains is supported by the images seen with the single MFab decoration, the 83-7 MFab epitope in the cys-rich region being spatially mapped roughly half-way up the side-edge of the ectodomain plates, and the 83-14 and 18-44 MFab epitopes (connecting domain and β-chain insert domain, respectively) being mapped near the base of the plates. Our preference is for a single a-b¢ monomer to occupy a single plate, although the possibility of a single monomer straddling the two plates of protein density cannot be discounted.

The more complex images involving co-binding of two, and even more so of all three, MFabs to each monomer of the ectodomain dimer are not easily interpretable with respect to relative domain arrangements within the monomer at present, not least of all because of the difficulty of finding conditions of negative staining that will simultaneously maintain the integrity of the Fab binding while highlighting recognisable and reproducible details of the internal structure of the dimeric IR ectodomain.

The data presented here demonstrate the ability of single-molecule imaging to give an initial insight into the topology of multidomain structures such as the ectodomain of hIR, and the value of combining this technique with that of either single or multiple monoclonal Fab attachment per monomer as a potential means of epitope, and domain, mapping of the structure. By imaging Fab complexes of other members of the family, such as hIGF-1R ectodomain, and combining available sequence-mapped epitope information with that presented here, a more comprehensive understanding of domain arrangements within the IR family ectodomains should be forthcoming.

EXAMPLE 5

Structure-Based Design of Ligands for the IGF Receptor as Potential Inhibitors of IGF Binding

The structure of IGF receptor can be considered as a filter or screen to design, or evaluate, potential ligands for the receptor. Those skilled in the art can use a number of well known methods for de novo ligand design, such as GRID, GREEN, HSITE, MCSS, HINT, BUCKETS, CLIX, LUDI, CAVEAT, SPLICE, HOOK, NEWLEAD, PRO_LIGAND, ELANA, LEGEND, GenStar, GrowMol, GROW, GEMINI, GroupBuild, SPROUT, and LEAPFROG, to generate potential agonists or antagonists for IGF-1R. In addition, the IGF-1R structure may be used as a query for database searches for potential ligands. The databases searched may be existing eg ACD, Cambridge Crystallographic, NCI, or virtual. Virtual databases, which contain very large numbers (currently up to 10^{12}) of chemically reasonable structures, may be generated by those skilled in the art using techniques such as DBMaker, ChemSpace, TRIAD and ILIAD.

The IGFR structure contains a number of sites into which putative ligands may bind. Search strategies known to those skilled in the art may be used to identify putative ligands for these sites. Examples of two suitable search strategies are described below:

(i) *Database Search*

The properties of key parts of the putative site may be used as a database search query. For example, the Unity 2.x database software may be used. A flexible 3D search can be run in which a "directed tweak" algorithm is used to find low energy conformations of potential ligands which satisfy the query.

(ii) *De novo design of ligands*

The Leapfrog algorithm as incorporated in the software package, Sybyl version 6.4.2 (Tripos Associates, St Louis), may be used to design potential ligands for IGF-1R sites. The coordinates of residues around the site may be taken from the x-ray structure, hydrogens and charges (Kollman all atom dictionary charges) added. From the size, shape and properties of the site, a number of potential ligands may be proposed. Leapfrog may be used to optimize the conformation of ligands and position on the site, to rank the likely strength of binding interactions with IGF-1R, and to suggest modifications to the structures which would have enhanced binding.

It is also possible to design ligands capable of interacting with more than one site. One way in which this may be done is by attaching flexible linkers to ligands designed for specific sites so as to join them. The linkers may be attached in such a way that they do not disrupt the binding to individual sites.

All references cited above are incorporated herein in their entirety by reference.

It will be appreciated by persons skilled in the art that numerous variations and/or modifications may be made to the invention as shown in the specific embodiments without departing from the spirit or scope of the invention as broadly described. The present embodiments are, therefore, to be considered in all respects as illustrative and not restrictive.